

# *Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource*

Article

Published Version

Romo, A., Amezcua, J. and Probst, O. (2011) Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource. *Journal of Renewable and Sustainable Energy*, 3 (2). 023105. ISSN 1941-7012 doi: <https://doi.org/10.1063/1.3574447> Available at <https://centaur.reading.ac.uk/28606/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1063/1.3574447>

Publisher: American Institute of Physics

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource

Alejandro Romo Perea,<sup>1,a)</sup> Javier Amezcua,<sup>2,b)</sup> and Oliver Probst<sup>1,c)</sup>

<sup>1</sup>*Department of Physics, Instituto Tecnológico y de Estudios Superiores de Monterrey, CP 64849, Monterrey, Nuevo León, Mexico*

<sup>2</sup>*Department of Atmospheric and Oceanic Sciences, University of Maryland, College Park, Maryland 20742-2425, USA*

(Received 7 October 2010; accepted 13 March 2011; published online 4 April 2011)

The estimation of the long-term wind resource at a prospective site based on a relatively short on-site measurement campaign is an indispensable task in the development of a commercial wind farm. The typical industry approach is based on the measure-correlate-predict (MCP) method where a relational model between the site wind velocity data and the data obtained from a suitable reference site is built from concurrent records. In a subsequent step, a long-term prediction for the prospective site is obtained from a combination of the relational model and the historic reference data. In the present paper, a systematic study is presented where three new MCP models, together with two published reference models (a simple linear regression and the variance ratio method), have been evaluated based on concurrent synthetic wind speed time series for two sites, simulating the prospective and the reference site. The synthetic method has the advantage of generating time series with the desired statistical properties, including Weibull scale and shape factors, required to evaluate the five methods under all plausible conditions. In this work, first a systematic discussion of the statistical fundamentals behind MCP methods is provided and three new models, one based on a nonlinear regression and two (termed kernel methods) derived from the use of conditional probability density functions, are proposed. All models are evaluated by using five metrics under a wide range of values of the correlation coefficient, the Weibull scale, and the Weibull shape factor. Only one of all models, a kernel method based on bivariate Weibull probability functions, is capable of accurately predicting all performance metrics studied. © 2011 American Institute of Physics. [doi:10.1063/1.3574447]

### I. INTRODUCTION

Typical wind resource assessment campaigns for commercial wind farm development last anywhere from one to three years, with important decisions to be taken often only after several months. There is therefore an obvious need for a prediction of the performance of a planned wind farm during its expected lifetime (20 yr or more). Such an assessment is an important part of the wind farm financing process. While the measurement campaign may correspond to an untypically high or low period, correlations with nearby reference stations should help detect such trends and provide a corrected long-term estimate of the wind speed at the development site and its interannual variations. Moreover, since the power output of wind turbine depends on the wind speed in

---

<sup>a)</sup>Electronic mail: romo.alex@itesm.mx. URL: <http://fisica.mty.itesm.mx/fisica/WEG/>.

<sup>b)</sup>Electronic mail: jamezcua@atmos.umd.edu.

<sup>c)</sup>Author to whom correspondence should be addressed. Electronic mail: oprobst@itesm.mx. URL: <http://fisica.mty.itesm.mx/fisica/WEG/>.

a nonlinear way, it is not enough to have an estimation only of the first moment (expected value) of the wind speed, but also of higher order moments (particularly the variance). If possible, it is desirable to estimate the actual distribution of the wind speed values.

The typical industry-standard approach to this problem is called the measure-correlate-predict (MCP) method. Some of the MCP approaches typically used in the wind energy industry have been reviewed by Rogers *et al.*<sup>1</sup> and Andersen.<sup>2</sup> Many models rely on some version of a linear fit between the reference and the site data for the concurrent period, with different approaches as to parameter estimation (see Ref. 2 for a detailed discussion). While often a standard least-squares fit is used to determine the fit parameters,<sup>3</sup> both for models with arbitrary vertical axis intercept and forced-zero intercept,<sup>2</sup> more generalized approaches consider versions of orthogonal regression where the distance between the measurements and the fitted line is minimized. This method is believed to be more suitable than a standard least-squares fit if both sets of measurements have an associated error,<sup>4</sup> as opposed to standard regression where only errors in the dependent variable (i.e., the site wind speed) are assumed to occur. Another distinction refers to the treatment of wind direction; some authors included the wind direction as an additional variable in the regression analysis (see, e.g., Ref. 5), whereas others<sup>6</sup> considered a linear vector relationship between the site and reference wind velocities expressed in orthogonal coordinates. Generally, however, the wind speed observations are binned by their wind direction values and a regression analysis is performed for each of the wind direction bins. Since site and reference wind direction values do not always pertain to the same wind direction bin, a decision has to be taken as to whether the site or the reference wind direction is used for binning; generally, the reference wind direction is used for that purpose. In the case where significant discrepancies between site and reference wind direction readings occur, this approach may not be appropriate; an alternative has been proposed by Woods and Watson,<sup>7</sup> where the predicted wind speeds pertaining to a wind direction sector  $i$  are obtained from a weighted average of the linear regression relations for sectors  $j$  correlated with the sector  $i$ . The weights are obtained from a matrix containing the occurrences of wind events with directions in the reference/site pair  $(i, j)$  subject to some threshold conditions.

While linear regression methods can, in principle, be generalized to nonlinear methods, assuming, e.g., a power law relationship between the reference and the site wind speed<sup>3</sup> or using second-order polynomial fits,<sup>5,8</sup> it is doubtful if such nonlinear regressions will provide a better long-term prediction of the wind resource at the site if they are not backed by an underlying statistical or physical model. An alternative is to go beyond the regression approach and retain some of the correlation structure between the two sites in the model. Such an approach has been formulated by Mortimer,<sup>9</sup> in which the concurrent data at the two sites are binned by the wind direction sector and wind speed at the reference site. Within each bin the ratios of the concurrent target and reference wind speeds are calculated. The predicted wind speeds in each wind direction bin are then obtained from a linear relationship  $\hat{y} = (r + e)x$ , where  $r$  is the average wind speed ratio in the respective wind speed and wind direction bin and  $e$  is a random variable with a triangular distribution with the appropriate standard deviation. A similar approach has been proposed by Andersen;<sup>2</sup> in their review, they discussed a prediction method based on the joint wind speed distribution for the reference-site pair for the concurrent period.

Systematic assessments of the performance of different MCP have been conducted both by Rogers *et al.*<sup>1</sup> and Andersen.<sup>2</sup> Andersen<sup>2</sup> focused mostly on the prediction accuracy of the average long-term wind speed, as well as the corresponding turbine yield for a selected wind turbine; they also presented some results as to the dependence of bias errors for wind speed and turbine yield on concurrency period and correlation coefficient. Rogers *et al.*<sup>1</sup> provided a systematic evaluation of four methods reported in literature as a function of the concurrency period. Their evaluation regarding the performance parameters assessed is more comprehensive since they assess not only the predicted versus the long-term average wind speed but also additional parameters representative of the wind speed distributions such as the standard deviation of the hourly values, as well as the Weibull shape and scale parameters. Rogers *et al.* also avoided the influence of seasonal variations of the results by averaging the performance metrics obtained from predictions with different nonoverlapping concurrency intervals of the same length.

While all studies discussed above provide useful insights, they all rely on specific sets of wind data, generally obtained from automatic surface observation stations (ASOSs). Remotely operated stations generally require a great deal of quality control and data preprocessing, and the raw data therefore have to be used with care. Although many quality control issues can be detected and appropriately accounted for, such as data gaps, abnormally high or low values, abnormally small temporal variations, outliers, etc. (see Ref. 10 for a recent analysis of ASOSs in Spain), other problems, such as anemometer soiling and varying exposure conditions, are much more difficult to detect. Another issue is related to instrument change, occurring both in an occasional manner at specific stations and during systematic replacements efforts, such as the recent campaign by the National Oceanographic and Atmospheric Administration (NOAA) in the United States, where cup anemometers have been replaced by sonic anemometers in a continuous effort over the past few years (see Ref. 11 for a discussion of the possible impact on MCP assessments). All of the phenomena mentioned above have a significant impact on the long-term prospection of the wind resource. This makes it difficult to assess a specific set of MCP methods without significant effects from the convolution with the data sets used; in other words, such assessments are case studies rather than general evaluations of the methods under study.

In the present work, synthetic time series with a controlled correlation structure and the desired wind speed distribution parameters have been used in order to conduct a systematic assessment of five different MCP methods. Three of these models are new and will be discussed below; the fourth and fifth models, a simple linear regression approach, as well as the variance ratio (VR) method put forward by Rogers *et al.*<sup>1</sup> have been used for reference purposes. The new MCP models in this work have been designed after a thorough analysis of the underlying bivariate probability density functions of two correlated wind speed time series, as opposed to *ad hoc* approaches often used in the MCP analyses. The first new model is a nonlinear regression model derived from a bivariate Weibull probability density function (*pdf*), whereas the second and third models are based on conditional probability density functions rather than regression curves and use bivariate Weibull and bivariate normal pdf's, respectively.

To evaluate these five methods, the metrics proposed by Rogers *et al.*<sup>1</sup> are used. The effect of concurrency period and cross-correlation coefficient on the performance metrics has been assessed. Additionally, advantage is taken of the fact that the input wind speed distributions of the concurrent time series can be manipulated by varying the Weibull shape parameters in a systematic manner.

## II. MCP MODELS STUDIED

### A. General MCP approach

In the widest sense, MCP methods address the problem of estimating the statistics of the long-term wind speed distribution of a target site  $v_{\text{site}}$  by using the available information from a reference site  $v_{\text{ref}}$ . The reference and target site wind speeds can be identified with random variables  $X$  and  $Y$ , which assume values of  $x$  and  $y$ , respectively.

The correlation procedure is a first step after which one performs a regression analysis where the wind speed of the target site is considered to be the dependent variable and the wind speed at the reference site the independent variable. In general, one can express a relationship of the type,

$$\hat{y} = g(x). \quad (1)$$

In Eq. (1),  $\hat{y}$  refers to the *predicted* or *estimated* wind speed at the site based on an observation of the reference wind speed  $x$ . One basic assumption is that Eq. (1) correctly describes the long-term relationship between the two random variables and not only during the correlation period, which is limited.

A deeper objective, which requires a more insightful analysis of the underlying statistical behavior of the process, is to achieve an estimation of the probability density function  $f_{Y,\text{long}}(y; \theta_Y)$ , which best describes the long-term wind speed distribution at a given location and height above ground level based on the knowledge of the density in the reference site  $f_{X,\text{long}}(x; \theta_X)$ .

$\theta$  is a vector of parameters for each one of the distributions; the number and type of parameters here depend on the particular distributions. The knowledge of  $f_{Y,\text{long}}(y; \theta_Y)$  allows to calculate the average wind speed and power density and, most importantly, the average energy yield of a given wind turbine. Additionally, intra- or interannual fluctuations of the wind resource at the prospective site need to be studied, although the present study does not consider such variations.

Due to the conservation of probability, we have

$$f_{Y,\text{long}}(y)dy = f_{X,\text{long}}(x)dx, \quad (2)$$

and therefore,

$$f_{Y,\text{long}}(y)dy = f_{X,\text{long}}[g^{-1}(y)] \frac{1}{dg/dx}, \quad (3)$$

where we have assumed that relationship (1) can be inverted. As mentioned above, the regression analysis outlined above can be generalized easily by first sorting the two wind speed data sets into wind direction categories and performing the regression analysis for each wind direction sector separately.

## B. Linear MCP models

The simple linear regression (SLR)—also known as straight line regression—is the first choice when establishing a regression between two sets of measurements. This often includes time series, a category to which wind speed measurements belong. In the most orthodox statistical sense, the SLR is not completely suited for this type of data. This is because for a SLR to be valid, some requirements must be fulfilled; one of them is the independence among the different values of each variable, i.e., the values of  $X$  must be mutually independent and so must the values of  $Y$ . Wind speed measurements—as most atmospheric data—contain a temporal autocovariance structure; nonetheless, SLR can be used after certain considerations. First, the biggest problems of using SLR would arise if we were interested in the exact evolution of the process throughout time. In our case, we are interested in the summarized characteristics for a certain period of time without considering the actual exact order in which the measurements occurred. On the other hand, it is important to consider the sampling frequency and the total sample time in comparison with (a), depending on the strength of the cycles (diurnal, annual), one must avoid sampling just one portion of the cycle or limit the procedure to just some part of the cycle. For (b), it is important to know about the decorrelation times of the atmosphere. For the surface layer, the most important structures will occur within the turbulent regime. These are, however, of no concern since we are dealing with hourly means. Moreover, working with a period longer than the typical decorrelation times of the boundary layer (2–6 days) will avoid related issues.

Having explained the previous caveats, the SLR will be discussed. In the predicted value  $\hat{y}$  of the dependent random variable  $Y$  is obtained from a given value of the independent random variable  $X$  through the following linear transformation:

$$\hat{y} = \mu_y + \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - \mu_x), \quad (4)$$

where  $\mu_x$  and  $\mu_y$  are the average values of the  $x$  and  $y$  variables,  $\sigma_x$  and  $\sigma_y$  are the corresponding standard deviations, and  $\rho_{xy}$  is the correlation coefficient. In general, the regression equation is written as  $\hat{y} = \alpha_0 + \alpha_1 x$  and the estimator for the coefficients is obtained directly via an original least-squares analysis rather than estimating the statistics for each variable and the correlation.

As mentioned earlier, in the context of wind energy prospection  $X$  is to be interpreted as the wind speed at a suitable reference station with a long-term record, whereas  $Y$  is the wind speed at a prospective site where only a limited amount of wind speed information (typically a year) has been logged. It is instructive to calculate the standard deviation of the predicted  $y$  values as

$$\sigma_{\hat{y}} = \rho_{xy} \sigma_y. \quad (5)$$

It is conspicuous from Eq. (5) that in a linear regression the standard deviation of the predicted values is generally smaller than that of the original data since the correlation coefficient is  $\rho_{xy} \leq 1$ .

### C. The variance ratio method

In wind resource assessment, we are interested in calculating the average wind power in the target site. Wind power depends on the third power of  $v$ , i.e.,  $p = av^3$ . Considering  $v$  as a random variable, our quantity of interest is then  $E[p] = E[v^3] = \mu^3 + \sigma^2(3\mu + \gamma\sigma)$ . In this expression,  $\mu$  is the expected value,  $\sigma^2$  is the variance, and  $\gamma$  is the asymmetry coefficient of the distribution of  $v$ . We must notice that besides an estimation of the expected value of our variable  $v_{\text{site}} = y$ , we require the estimation of higher order moments, particularly the variance (the asymmetry coefficient is usually small).

With the SLR discussed earlier, the standard deviation of the predicted wind speed values of the target site constitutes an underestimation of the actual standard deviation of wind speed values measured in the target site. This is a detrimental issue since, as stated before, the mean power production of a given wind turbine will depend on measures of fluctuations. Rogers *et al.*<sup>1</sup> observed this drawback and proposed a new correlation method dubbed the VR method; in this approach, the predicted variable is obtained from the independent variable through the following relationship:

$$\hat{y} = \mu_x + \frac{\sigma_y}{\sigma_x}(x - \mu_x). \quad (6)$$

Equation (6) can be viewed as a special case of Eq. (4) with the correlation coefficient set to 1. By construction, the average value of the predicted  $y$  values is equal to the mean of the measured values, just like in the SLR. Moreover, however, the standard deviation is also the same, as it can be seen from Eq. (5) with  $\rho_{xy} = 1$ . In case the functional relationship between the wind speeds at the two sites remains the same over time, the VR method should be able to accurately predict both the first and the second moment of the site wind speed distribution (based on the concurrent observations at the reference site), whereas the simple linear regression will only correctly predict the first moment. At this moment, it is less clear how either method will perform with respect to higher moments and parameters of interest related to wind speed distributions, such as Weibull parameters and energy density. These issues will be extensively addressed in the results section.

### D. Linear regression derived from bivariate normal joint distribution

In order to design a MCP method that will accurately predict all parameters typically of interest for wind energy applications, it is instructive to first analyze the standard methods, such as the SLR, from a more statistical viewpoint. To that end, it is first observed that correlated pairs of variables can be conveniently modeled by a bivariate joint probability density function (*bi-pdf*)  $f(x, y)$ . Often, it is assumed that the pdf is bivariate normal, i.e.,

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right\}. \quad (7)$$

Marginal probability density functions (*m-pdf*), i.e., those pertaining to either variable ( $x$  or  $y$ ) independently of the other, are obtained by integrating over the variable which is not of interest. Marginal pdf's of a bivariate normal pdf are again normal.

Another important element is the conditional density function (*c-pdf*)  $f_Y(y|X=x)$ . This is the



probability density of the dependent random variable  $Y$  given a certain value ( $x$ ) of the independent random variable  $X$ . The c-pdf is defined as

$$f_Y(y|X=x) = \frac{f_{XY}(x,y)}{f_X(x)}, \quad (8)$$

where  $f_{XY}(x,y)$  is the bivariate joint pdf and  $f_X(x)$  is the marginal pdf for the independent variable, i.e., the reference wind speed. It is easy to show that  $f_Y(y|X=x)$  is normalized to 1, i.e., it can be interpreted as a marginal probability density function for the variable of interest (the site wind speed), given a certain observation  $x$  of the reference wind speed. For a bivariate normal joint pdf, both conditional pdf's are again normal.

We will now discuss the regression function of  $Y$  on  $X$  defined by

$$\tilde{\mu}_{y \cdot x} = E[Y|X=x] = \int_{-\infty}^{\infty} y f_Y(y|X=x) dy. \quad (9)$$

By definition,  $\tilde{\mu}_{y \cdot x}$  is the expected value of the dependent variable (site wind speed), given a certain observation  $x$  of the reference wind speed, i.e., a regression value. Moreover, in the case of a conditional pdf derived from a bivariate normal joint pdf, it can be shown that

$$\tilde{\mu}_{y \cdot x} = \mu_{y \cdot x} = \mu_y + \frac{\rho_{xy}\sigma_y}{\sigma_x}(x - \mu_x). \quad (10)$$

It is conspicuous from Eq. (10) that the  $X$ - $Y$  regression derived from the conditional pdf yields the familiar linear regression stated in Eq. (4). In other words, the simple linear regression arises naturally as an expected value of the conditional pdf for a bivariate joint normal distribution of the random variables  $X$  (reference wind speed) and  $Y$  (site wind speed). As a side note, and without further derivations for the sake of brevity, it can be stated that the conditional pdf is centered at the regression value of  $Y$  for a given value of  $X=x$ , and that the standard deviation of  $Y$  is independent of  $x$  [see Eq. (10)].

## E. The Weibull regression approach

While it has been discussed that a linear regression emerges naturally if a bivariate joint-normal distribution of the variables  $X$  and  $Y$  is assumed, in the case of wind speed prospection this approach is questionable. A bivariate joint-normal distribution implies normal marginal distributions. Hourly wind speed averages, the ones typically used in wind energy prospection, do not follow normal distributions but are described rather by Weibull distributions.<sup>12</sup> It is therefore plausible to look for a bivariate pdf whose marginal probability density functions are Weibull. This type of bivariate pdf's have been proposed by several authors.<sup>13-17</sup> Here, the following cumulative pdf (or *cdf*) proposed by Johnson *et al.*<sup>18</sup> will be used:

$$F(x,y) = 1 - \exp \left\{ - \left[ \left( \frac{x}{\lambda_x} \right)^{\kappa_x/\delta} + \left( \frac{y}{\lambda_y} \right)^{\kappa_y/\delta} \right]^\delta \right\}, \quad 0 < \delta \leq 1. \quad (11)$$

It can be shown that the marginal pdf's are Weibull, where  $\lambda_x, \lambda_y$  are the scale and  $\kappa_x, \kappa_y$  are the shape factors of the distributions for  $X$  and  $Y$ , respectively.  $\delta$  is a parameter controlling the degree of association between the two variables.

It is important to mention that the linear correlation coefficient  $\rho_{xy}$  is, however, a nonlinear monotonical function of  $\kappa_x, \kappa_y$ , and  $\delta$ . That is,  $\rho_{xy}$  is independent of the scale factors of the marginal distributions. Figure 1 illustrates the relationship between the two parameters for the case of equal shape factors [Fig. 1(a)] and the case where one of the shape factors doubles the value of the other [Fig. 1(b)]. It can be seen from Fig. 1 that for very small values of the shape factor  $\rho_{xy}$  varies very strongly with  $\delta$ , so that in these cases  $\delta$  may not be an appropriate parameter for the control of the correlation coefficient. For the typical range of shape factors encountered in wind



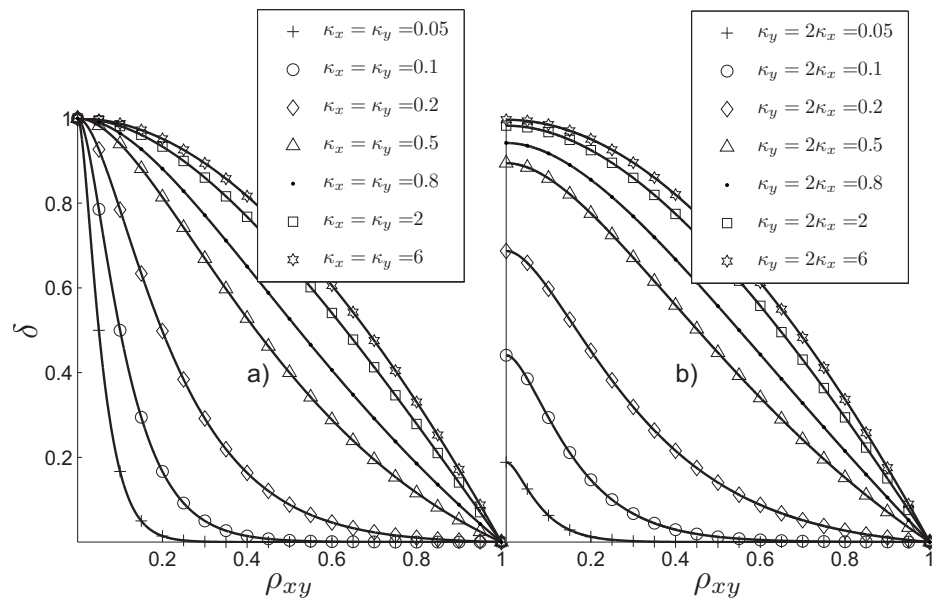


FIG. 1. Relationship between  $\delta$  and  $\rho$  for several values of  $\kappa_x = \kappa_y$ . In the left plot,  $\kappa_x = \kappa_y$  takes values in the range  $[0.05, 6]$ . The right plot depicts some cases where  $\kappa_x = \kappa_y/2$ .

energy prospection, however, i.e.,  $\kappa_x, \kappa_y = 1, \dots, 3.5$ , a nearly linear relationship between the two parameters exists, so it is easy to specify the correlation coefficient by specifying  $\delta$ .

After these introductory remarks, it is possible to calculate the conditional density probability function  $f_{XY}^{\text{Weibull}}(y|X=x)$  and the regression function  $E^{\text{Weibull}}(Y|X=x)$ . As expected from the additional parameters present in the bivariate Weibull joint pdf, both the conditional pdf and the regression function have a more complex structure compared to the bivariate-normal case. In Fig. 2, the regression function between  $X$  and  $Y$  has been plotted for the case of different shape factor ratios (different symbols in each subgraph) and correlation coefficients. The shape factors used cover the typical range encountered in wind speed measurements; the scale factor was set to 7 m/s for all cases and both sites. It can be seen from the figure that all curves intersect at the point  $(\lambda_x, \lambda_y)$ ; this observation still holds if the scale factors are different (not shown). Interestingly, for the highest value of the correlation coefficient (0.95), a near-linear relationship is observed, although even in this case a marked dependence of the shape factor ratio can be seen. At lower correlation coefficients, the regression curves generally show a convex shape, which is analogous to the situation of the linear regression where a low correlation leads to a more horizontal curve and therefore a higher vertical axis intercept [cf. Eq. (4)].

## F. Approaches based on conditional probability density functions

A general aspect shared by MCP approaches based on regression models is the fact that the predicted or *backcast* values of the dependent variable (to be identified with the wind speed data at the prospective site) have a strict functional relationship with the independent values [given by Eq. (1)], which is of course unrealistic. As long as only the average value (the first moment) for a given time interval is of interest, this may be tolerable; but if higher moments and functions thereof are required, a regression approach may not be appropriate.

A possible solution to this problem is an approach that retains information about the joint probabilistic structure of the two data sets. As discussed before, the ultimate objective of the MCP process should be to construct a marginal probability density function that accurately describes the long-term wind speed distribution at the prospective site. While the approach expressed by Eq. (3) implies a (linear or nonlinear) transformation of the long-term reference pdf, the more general

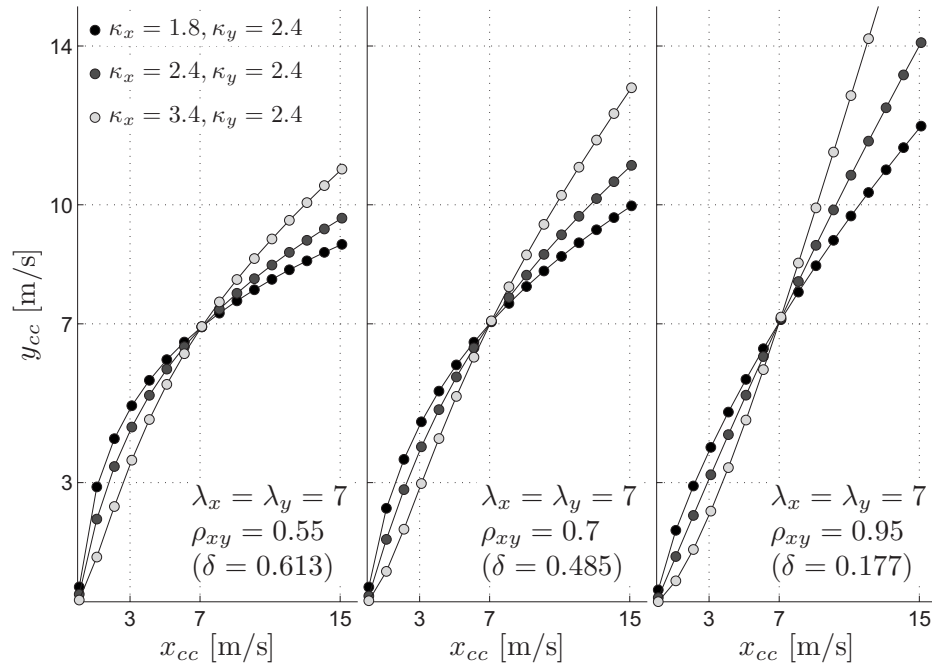


FIG. 2. Regression curves obtained with a bivariate-Weibull joint distribution for the wind speed for a reference/target site pair for different values of the cross-correlation coefficient  $\rho_{xy}$ : (a)  $\rho_{xy}=0.55$ , (b)  $\rho_{xy}=0.7$ , and (c)  $\rho_{xy}=0.95$ .

approach put forward in this paper relies on obtaining a full conditional pdf during the concurrent period rather than on performing a regression (i.e., the expectation value of the dependent variable calculated from the c-pdf).

Consider the conditional probability density function given by

$$f_{XY}^{(\text{conc})}(y|X=x) = \frac{f_{XY}^{(\text{conc})}(x,y)}{f_X^{(\text{conc})}(x)}, \quad (12)$$

determined from concurrent reference and site wind speed data, where  $f_{XY}^{(\text{conc})}(x,y)$  and  $f_X^{(\text{conc})}(x)$  are the bivariate pdf and marginal pdf for the reference data, respectively, and the superscript “conc” refers to the concurrent period. Clearly, the marginal pdf for the site wind speed data can be obtained by multiplying the c-pdf with the marginal pdf of the reference data and integrating over the reference wind speed variable,

$$f_Y^{(\text{conc})}(x) = \int f_{XY}^{(\text{conc})}(y|X=x) f_X^{(\text{conc})}(x) dx. \quad (13)$$

While Eq. (13) is exact, the question arises how to model the conditional pdf for observation periods where no concurrent information is available. The hypothesis put forward in the present work is that the conditional pdf can be taken to be constant over the period of interest, i.e.,

$$f_Y(x) = \int f_{XY}^{(\text{conc})}(y|X=x) f_X(x) dx \quad (14)$$

for any observation period where quality reference data are available.

In order to implement the approach summarized by Eqs. (12)–(14), a suitable expression for the c-pdf has to be found. One approach, the one put forward in the present work, is to determine the bivariate Weibull joint distribution function [obtained by deriving Eq. (11)] that best fits the concurrent data and then calculate  $f_{XY}^{(\text{conc})}(y|X=x)$  (see Fig. 3 for an example) from Eq. (12) using

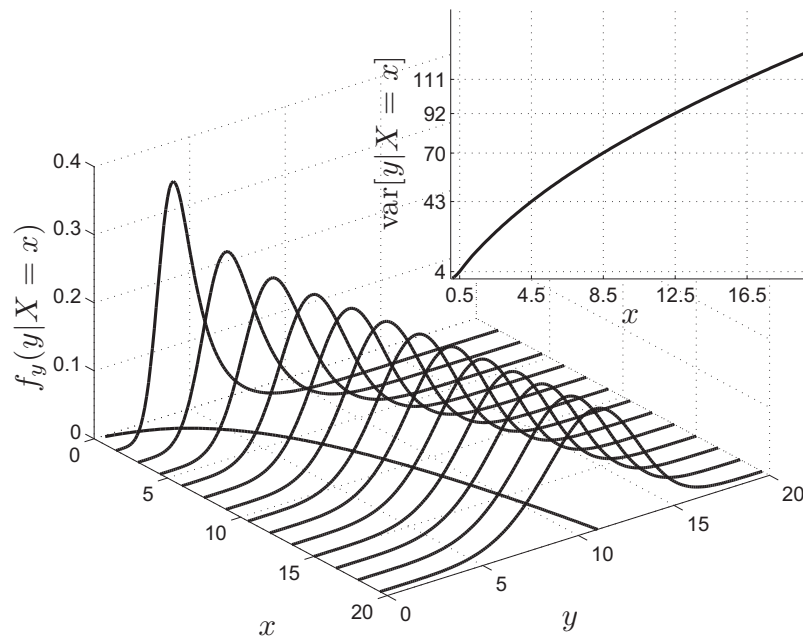


FIG. 3. Bivariate conditional Weibull pdf for several values of  $X=x$  ( $\lambda_x=6.5$ ,  $\kappa_x=1.8$ ,  $\lambda_y=7.4$ ,  $\kappa_y=2.8$  and  $\delta=0.5$ ). The expected value of the conditional pdf, or regression function, is shown in the  $x$ - $y$  plane. The respective variance function is shown in the plot.

the measured marginal pdf of the reference site. As an alternative, the bivariate normal distribution function that best fits the data could be used. Such an approach was proposed recently by Chan *et al.*,<sup>19</sup> albeit in the context of hourly forecasts. The authors referred to their method as a kernel method,<sup>19</sup> with Gaussian kernels used in their work. We will refer to the method using a bivariate Weibull pdf as *Wpdf*, where the method using the bivariate normal distribution will be termed *SLRpdf* due to the close relation born by the simple linear regression with binormal distributions (Sec. II D).

### G. Parameter estimation

In the case of the SLR, parameter estimation is straightforward; the two fitted parameters of the regression equation (slope and intercept) depend on the mean values and standard deviations of the two data sets and their correlation coefficient. It is a standard procedure to directly estimate these two parameters ( $\alpha_0, \alpha_1$ ) using original least-squares. To assess the certainty in the estimators, it is possible to find confidence intervals for both. A confidence interval at the level  $(1-\gamma)$  for the estimator  $\alpha_i$  has the form,  $\alpha_i \in [\hat{\alpha}_i \pm s_{\hat{\alpha}_i} t_{n-2}(\gamma/2)]$ , where  $t_{n-2}(\gamma/2)$  is the  $1-\gamma/2$ th percentile of the  $t$  distribution with  $n-2$  degrees of freedom. On the other hand,  $s_{\hat{\alpha}_i}$  is the standard deviation of the estimator  $\hat{\alpha}_i$  and it decreases with the sample size as  $s_{\hat{\alpha}_i} \propto (n-2)^{-1/2}$  (see, for example, Ref. 20 for the complete expression). Moreover, it is possible to find confidence bands for the regression line itself; the width of the bands depend on the specific value of  $x$  and are also proportional to  $(n-2)^{-1/2}$ .

In the case of the VR method, the situation is even simpler since the correlation coefficient has been set to 1. For the two new methods proposed, however, parameter estimation is a more complex procedure since it relies on a nonlinear fit of a bivariate Weibull pdf to the concurrent two data sets. For each site pair and concurrent period, a set of five parameters is obtained; these parameters fully specify the bivariate joint Weibull distribution and the conditional pdf is consequently specified.

To estimate the five free parameters of this bivariate pdf (reference scale, shape factors, site

scale, shape factors, and association parameter  $\delta$ ), we use maximum-likelihood estimation (MLE). In this procedure, one maximizes the natural logarithm of the likelihood function, explicitly,<sup>18</sup>

$$\begin{aligned} \log L = & n \log \left( \frac{\kappa_x}{\lambda_x} \right) + n \log \left( \frac{\kappa_y}{\lambda_y} \right) + \left( \frac{\lambda_x}{\delta} - 1 \right) \sum_{i=1}^n \log \left( \frac{x_i}{\lambda_x} \right) + \left( \frac{\lambda_y}{\delta} - 1 \right) \\ & \times \sum_{i=1}^n \log \left( \frac{y_i}{\lambda_y} \right) + (\delta - 2) \sum_{i=1}^n \log \left[ \left( \frac{x_i}{\lambda_x} \right)^{\kappa_x/\delta} + \left( \frac{y_i}{\lambda_y} \right)^{\kappa_y/\delta} \right] \sum_{i=1}^n \log \left[ \left[ \left( \frac{x_i}{\lambda_x} \right)^{\kappa_x/\delta} + \left( \frac{y_i}{\lambda_y} \right)^{\kappa_y/\delta} \right]^\delta \right. \\ & \left. + \frac{1}{\delta} - 1 \right] - \sum_{i=1}^n \left[ \left( \frac{x_i}{\lambda_x} \right)^{\kappa_x/\delta} + \left( \frac{y_i}{\lambda_y} \right)^{\kappa_y/\delta} \right]^\delta. \end{aligned} \quad (15)$$

Unlike in the SLR, in this case no analytical expression for the estimators can be found, so the maximization has to be performed numerically.

With respect to the accuracy of the estimators, we know that they are both efficient and consistent (properties of the MLE) and that whatever bias present will fade at the rate  $n^{-1/2}$ . Nonetheless, constructing confidence intervals is quite more complicated. Let  $\theta = [\kappa_x, \kappa_y, \lambda_x, \lambda_y, \delta]^T$ , i.e., the vector of parameters. Then, the vector of estimators  $\hat{\theta}$  has the following asymptotic multivariate normal distribution:  $\sqrt{n}(\hat{\theta} - \theta) \sim N(0, I^{-1})$ , where the covariance is just the inverse of the Fisher's information matrix for the bivariate Weibull distribution, which can be calculated as  $I(\theta) = E([\nabla_\theta \ln L(\theta)][\nabla_\theta \ln L(\theta)]^T)$ , the explicit form of this  $5 \times 5$  matrix can be found in Ref. 20. This matrix is also useful to construct asymptotic confidence intervals for the regression itself; a method to achieve this is described in Ref. 21.

For our synthetic experiments, since our sample size is close to  $10^6$ , we were able to estimate the parameters quite accurately. Nonetheless, the previous considerations are important, especially for smaller samples (e.g., less than  $10^3$ ), and should be investigated in further detail in subsequent works.

### III. GENERATION OF CORRELATED TIME SERIES WITH WEIBULL DISTRIBUTIONS

The general task described in this section is the generation of two correlated concurrent time series with two independent Weibull marginal distributions and independent autocorrelation lengths. This task is divided into two sections, the generation of normally distributed time series with an established correlation structure (auto- and cross-correlation), which will be described in Sec. III A, and the transformation to series with marginal Weibull distribution (to be discussed in Sec. III B).

#### A. Generation of correlation structure

In order to generate the required two wind speed time series with an established correlation structure, a bivariate first-order vector autoregressive  $[\text{VAR}_2(1)]$  process is used. As usual, the deviations of two variables from their mean value at a time step  $t$ , grouped into the vector  $\tilde{\mathbf{Z}}_t$ , are calculated from the following recursion (see Ref. 22, Chap. 1, and Ref. 23, Chap. 14):

$$\tilde{\mathbf{Z}}_t = \Phi_1 \tilde{\mathbf{Z}}_{t-1} + \mathbf{a}_t, \quad (16)$$

where  $\tilde{\mathbf{Z}}_t = (\tilde{z}_{1t}, \tilde{z}_{2t})^T$  is a random  $(2 \times 1)$  vector of deviations from  $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ , i.e.,

$$\tilde{\mathbf{Z}}_t = \mathbf{Z}_t - \boldsymbol{\mu} = (\tilde{z}_{1t}, \dots, \tilde{z}_{kt})', \quad (17)$$

and  $(\mathbf{a})_t = (a_{1t}, a_{2t})^T$  is a bivariate normal white noise process on the interval  $[0, 1]$ . Hence, its covariance function is given by

$$E[\mathbf{a}_{t+l}\mathbf{a}_t'] = \begin{cases} \Sigma_a, & l = 0 \\ \mathbf{0}_{k \times k}, & l \neq 0, \end{cases} \quad (18)$$

where  $\Sigma_a$  is the covariance matrix,

$$\Sigma_a = \begin{pmatrix} \sigma_{1a}^2 & \rho\sigma_{1a}\sigma_{2a} \\ \rho\sigma_{1a}\sigma_{2a} & \sigma_{2a}^2 \end{pmatrix}, \quad (19)$$

and  $\Phi_1$  is a matrix of coefficients whose eigenvalues have all modulus  $< 1$  (see Ref. 22, Chap. 1).

## B. Shaping of univariate probability density functions

Once the two concurrent time series with its correlation structure have been defined, it is straightforward to transform each univariate time series separately into a time series with the desired distribution—in this case, a given Weibull probability function. Since the time series  $z_1$  and  $z_2$  are normally distributed random variables, the application of the normal cumulative probability density function (cdf) yields variables with a uniform distribution on the standard interval  $[0, 1]$ ,

$$U_1 = F_{N(0,1)}(z_1),$$

$$U_2 = F_{N(0,1)}(z_2).$$

In a second step, the inverse cdf of the desired distributions, in this case the inverse Weibull cdf, is applied to the uniformly distributed variables generating the following desired final time series:

$$X = T_1^{-1}(U_1),$$

$$Y = T_2^{-1}(U_2).$$

This process is known as inverse transform sampling or Smirnov transform and is based on the probability integral transform theorem.<sup>24</sup> It is noteworthy that this transformation process conserves monotonicity and the linear correlation structure defined in the first part of the time series generation process (Sec. III B). These methods have been used in the past.<sup>25</sup>

## C. Design of computational experiments

Now, the synthetic times series model (STSM) described in the last section are to be used to conduct a systematic assessment of the five MCP methods already mentioned. The procedure is summarized as follows:

- (1) A  $\text{VAR}_2(1)$  process of length  $n=87\,600$  is generated to resemble a 10-year period of hourly averages of wind speed for a reference-prospective site pair with Weibull parameters  $\lambda_x, \kappa_x$  and  $\lambda_y, \kappa_y$ , respectively. The correlation structure (see Sec. III A) is also specified.
- (2) A realization of the STSM setup before is produced.
- (3) The concurrent period to be used for the construction of the relational model is selected. The rest of the data corresponds to the historic data of reference and prospective sites, respectively.
- (4) A MCP analysis is carried out, i.e., a prediction is generated for the historic prospective site data by using the historic reference data and the relational model chosen.
- (5) Each model's performance is evaluated by using the metrics listed in Table I. These metrics are simply the ratio between the observed and the estimated quantity in question. Therefore, the closer to one is the metric of a model, the better its performance.
- (6) Steps (3)–(5) are repeated for different realizations of the STSM in step (1). This is done 25 times.

TABLE I. Metrics used to evaluate model's performance.

Metrics	
Mean	$m_{\mu} = \mu_{\hat{y}} / \mu_y$
Standard deviation	$m_{\sigma} = \sigma_{\hat{y}} / \sigma_y$
Scale factor	$m_{\lambda} = \lambda_{\hat{y}} / \lambda_y$
Shape factor	$m_{\kappa} = \kappa_{\hat{y}} / \kappa_y$
Energy density	$m_E = E_{\hat{y}} / E_y$

- (7) After running the 25 simulations of the last step, 25 results are obtained for each metric of step (5). The average of these values is taken for each metric. These are the final results of the performance of the model for the parameters defined set in step (1).
- (8) Steps (1)–(7) are repeated for other set of parameters. For this work,  $\lambda_x$ ,  $\kappa_x$ ,  $\lambda_y$ ,  $\kappa_y$ , and  $\rho_{xy}$  along with the concurrent time length (length of data used to construct the relational model) are varied in the ranges shown in Table II.

## IV. RESULTS AND DISCUSSION

### A. Influence of the concurrency period

A rapid convergence was observed for all performance metrics and input parameter combinations. In most cases, after 1000 time steps the performance metrics were within 1% of the convergence value. An example of this behavior is shown in Fig. 4 where the performance metrics for the mean (a) and the standard deviation (b) have been plotted as a function of the concurrency period for all five methods under study. The parameters used are as follows: Weibull scale factors  $\lambda_1 = \lambda_2 = 7.5$ , Weibull shape factors  $\kappa_1 = \kappa_2 = 3.0$ , the cross-correlation coefficient was 0.85, and the autocorrelation coefficient was 0.7. Note that four of the five methods correctly predict the long-term average wind speed (with a slight overprediction by the SLRpdf method), but only the VR and the Wpdf method correctly predict the long-term standard deviation of the wind speed values. The good convergence as a function of the concurrency period is plausible given the fact that no seasonal variations are considered in this study.

### B. Influence of the correlation coefficient

Only two of the methods analyzed, the simple linear (SLR) and the Weibull regression (WR), show a marked dependence of the performance metrics on the correlation coefficient  $\rho_{xy}$ . While for all methods the metric of the average wind speed is almost independent of  $\rho_{xy}$ , a strong dependence is observed for the SLR and WR methods in the case of the Weibull scale and shape factor, the standard deviation, and the energy density (Fig. 5). As seen in Fig. 5(a), in the case of the SLR and the WR method, the scale factor metric increases approximately linearly with  $\rho_{xy}$ , approaching unity for high values of the correlation coefficient. The shape factor [Fig. 5(b)], on the other hand, is overpredicted by the SLR and the WR method for low values of  $\rho_{xy}$ , with a convergence to unity for high correlations.

TABLE II. Space of parameters explored in this work.

Variable	Range
ccL	500–9500
$\lambda_x, \lambda_y$	4.5–14
$\kappa_x, \kappa_y$	1.3–5
$\rho_{xy}$	0.55–0.95

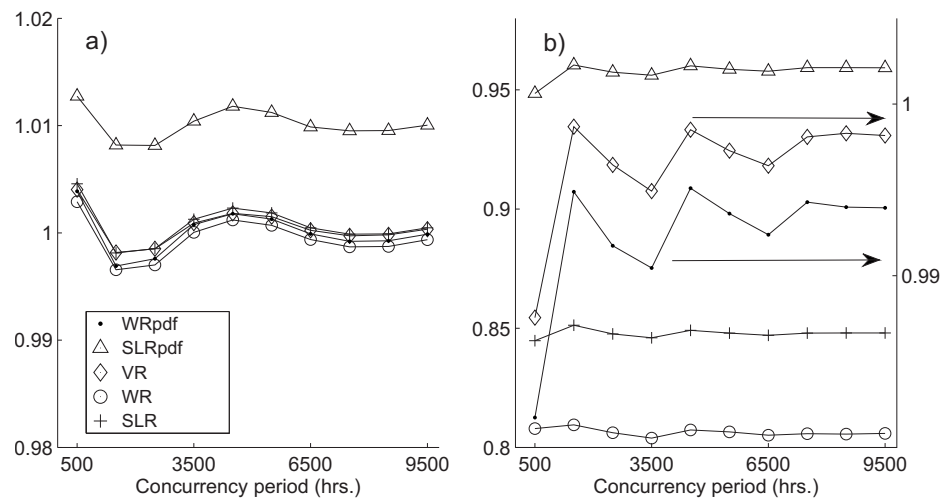


FIG. 4. Influence of the concurrency period on the performance metrics for the mean (a) and the standard deviation (b).

As we can see in Fig. 5(c), the ratio of the predicted versus the measured long-term standard deviation in the case of the simple linear regression increases linearly with  $\rho_{xy}$ , as expected from Eq. (5). A similar behavior is observed for the Weibull regression method. The SLRpdf method, on the other hand, shows practically no dependence on  $\rho_{xy}$ , but slightly underpredicts the long-term standard deviation.

### C. Influence of the shape factor ratio

Different Weibull shape factors are often encountered in MCP analyses, so it is interesting to study how MCP predictions vary with the ratio of the shape factors. For the calculations in this

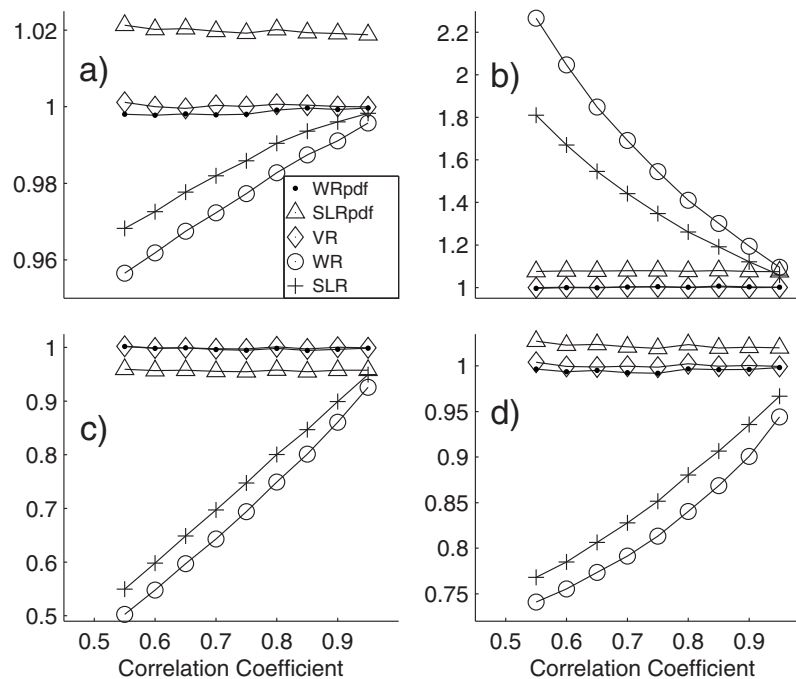


FIG. 5. Influence of the cross-correlation coefficient on the performance metrics for the (a) scale, (b) shape factor, (c) standard deviation, and (d) energy density.



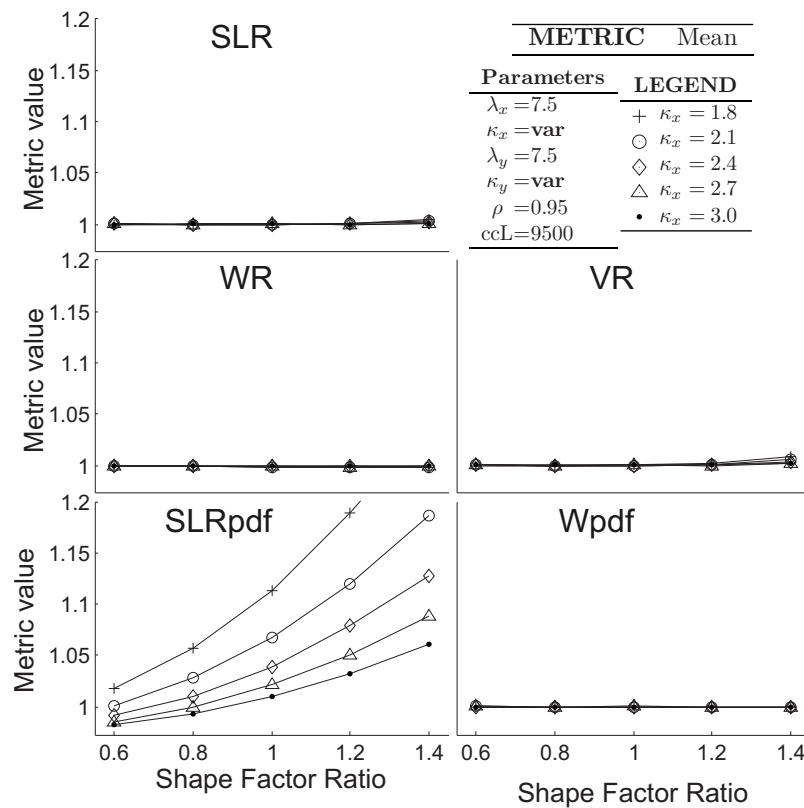


FIG. 6. Influence of the shape factor ratio on the predictions of the long-term average wind speed for the five methods studied.

section, both the site and the reference shape factors were varied in the range of 1.8–3.0, thereby covering the typical range encountered in practice (see Ref. 26, Chap. 4). The results (metrics) are plotted against the reference/site shape factor ratio, with different reference shape factor values indicated by different plot symbols. A concurrency period of 9500 time steps and a correlation coefficient of 0.95 were used in all cases.

It can be seen from Fig. 6 that all methods except for SLRpdf provide accurate predictions of the long-term average wind speed for all combinations of the site/reference shape factor values. The SLRpdf method, on the other hand, performs best in the case where  $\kappa_x = \kappa_y = 3.0$  (i.e.,  $\kappa_x = 3.0$  and  $\alpha = \kappa_x / \kappa_y = 1$ ), where the mean wind speed is correctly predicted. This finding comes at no surprise if we recall that Weibull distributions for a shape factor in the proximity of 3 closely resemble normal or Gaussian distributions (more specifically, skewness for a normal distribution is almost zero in the range 2.6 and 3.7). For all other shape factor combinations, however, the assumption of Gaussian kernels generally leads to a substantial overprediction of the average wind speed.

When the other metrics are analyzed, the differences between the different methods become far more substantial. A first example is the long-term standard deviation of the wind speed values (Fig. 7). The SLR consistently underpredicts the site shape factor (by some 5%–7% for the case of a correlation coefficient of 0.95) but does not show much variation of the shape factor ratio in the range studied and the results are essentially independent of the absolute value of the reference shape factor. For the highest value of the shape factor ratio (1.4), though, some dependence on the reference shape factor does show up. The new, nonlinear, regression (WR) also underpredicts the standard deviation (by some 6%–9%) and the prediction accuracy slightly worsens with increasing shape factor ratio. Also, some dependence on the individual values of the shape factors is ob-

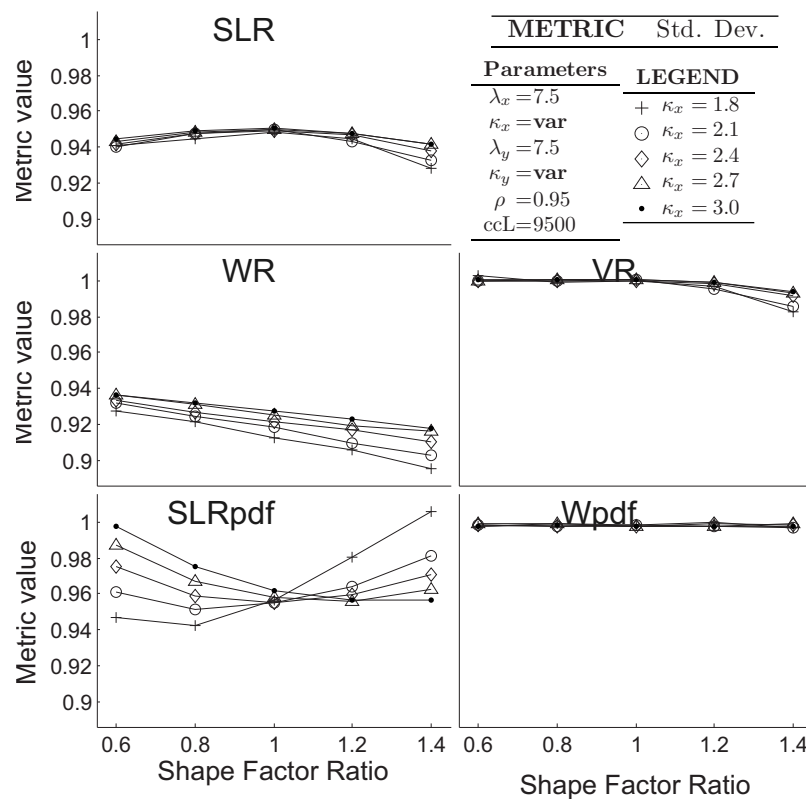


FIG. 7. Influence of the shape factor ratio on the predictions of the long-term standard deviation of the wind speed for the five methods studied.

served. The SLRpdf method generally provides a better prediction than both regression methods (underpredicting the standard deviation by some 0%–6%) but shows a stronger dependence on both the ratio and the absolute values of the shape factor.

As to the VR and the Weibull kernel method (Wpdf), both methods lead to excellent predictions of the standard deviation, but at higher values of the shape factor ratios (1.2 and 1.4) an advantage of the new method (Wpdf) over the VR approach becomes apparent. While the Wpdf very accurately predicts the standard deviation for all combinations of the shape factors, the VR shows some, though small, underprediction at high shape factor ratios.

As far as the prediction accuracy of the site shape factor is concerned, the effect of the (concurrent) shape factor ratio is quite substantial. As conspicuous from Fig. 8, the predictions of the Weibull methods (both WR and Wpdf) are independent of the shape factor ratio, while the three other methods show a significant dependence. Interestingly, the behavior of the SLR and the VR method is almost identical, with a significant underprediction of the true long-term shape factor at high concurrent shape factor ratios and a considerable dependence on the absolute value of the concurrent reference shape factor. As in other occasions, the SLRpdf method behaves unfavorably, overpredicting the true target site shape factor and showing a significant dependence on both concurrent shape factors. Only the Wpdf method provides a perfect prediction of the site shape factor in all occasions.

Regarding the scale factor metric, the situation is similar to the metric for the average wind speed, although the scale factor is a little less well predicted in the case of the SLR, WR, and even the VR method for shape factor ratios  $\alpha > 1$ . This behavior can be traced back to the relatively poor prediction of the target site shape factor using these methods for the case of  $\alpha > 1$ , remembering that for a Weibull distribution the mean  $\mu$ , the scale factor  $\lambda$ , and the shape factor  $\kappa$  are related by  $\mu = \lambda \Gamma(1 + 1/\kappa)$ , where  $\Gamma(\cdot)$  is the Gamma function (see Ref. 27, Sec. 3.3).

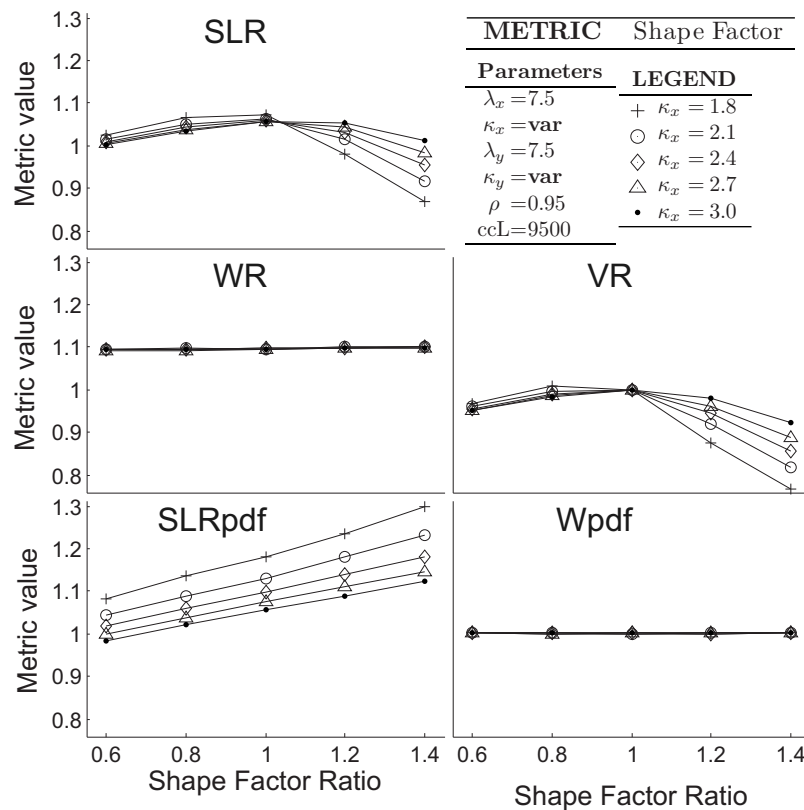


FIG. 8. Influence of the shape factor ratio on the predictions of the long-term shape factor of the wind speed distribution for the five methods studied.

It is interesting to analyze how the findings presented above translate into the predictions of the energy density. As shown in Fig. 9, only the Wpdf method provides a perfect prediction of the long-term energy density at the target site; all other methods show a significant deviation from the true value, particularly at high values of  $\alpha$ . While the SLR and the WR methods consistently underpredict the energy density, with an increasing discrepancy as  $\alpha$  increases, the SLRpdf method overpredicts the energy density in all cases, with an increasing overprediction for higher values of the shape factor ratio. The VR method behaves much better than either SLR, WR, or SLRpdf, with a good prediction for  $\alpha < 1$ . For  $\alpha > 1$ , however, the energy density is underpredicted.

#### D. Influence of the scale factor ratio

Extensive simulations were performed for variable scale factor ratios. As implied by its the name, the scale factors only change the scale used for each variable. A set of Weibull distributions with the same shape factor but different scale factors can be considered a scale family, and the members can be transformed to a *standard* member by a simple scaling conserving *equality in distribution*. The effect of changing the scale factor ratio should therefore be equivalent to changing units on one of the axes. We were able to verify that indeed none of the metrics depended on the scale factor ratio for any of the methods studied.

#### E. Summary and overall conclusions

In the present study, a systematic assessment of the five methods for the long-term prediction of the wind resource, known as MCP methods, has been conducted. The approaches studied can be divided into *regression* (SLR and WR) and *kernel* methods (Wpdf and SLpdf), where the latter is

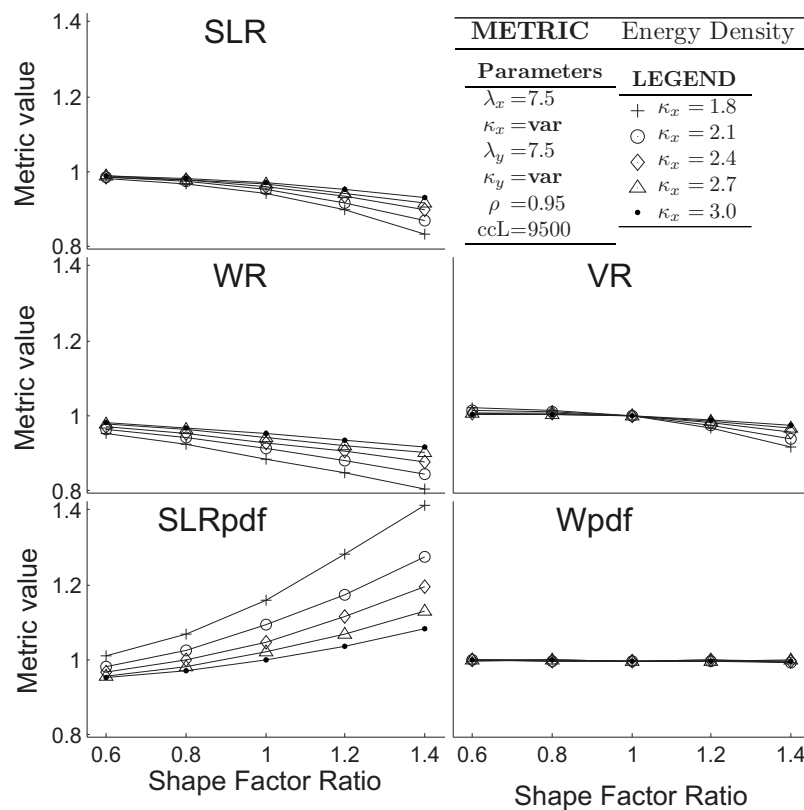


FIG. 9. Influence of the shape factor ratio on the predictions of the long-term energy density of the wind speed distribution for the five methods studied.

based on the concept of the conditional probability density function applied to the joint distribution of the reference and site wind speeds. The methods can also be classified by their parameter estimation scheme; while the SLR and SLRpdf methods use the simple linear regression parameters dependent on averages, standard deviations, and the cross-correlation coefficient (five parameters), both the WR and the Wpdf approaches are based on the five parameters of the bi-Weibullian probability density function which best approximates the data in a maximum-likelihood sense. The fifth method, reported by Rogers *et al.*<sup>1</sup> to outperform all other methods studied by them, was used as a reference and due to its simplicity. This model can neither be classified as a regression nor a kernel method.

As it can be seen from Table III, summarizing the performance of the different methods in a qualitative way, and Fig. 10, showing the range of the five performance metrics for a range of correlation coefficients  $\rho_{xy}$  ( $x \leq \rho_{xy} \leq y$ ) and shape factor ratios ( $x \leq \alpha \leq y$ ), the Wpdf method (proposed in this work) is the only technique that accurately predicts the long-term values of all five variables of interest under all parameter combinations. Its cousin, the SLRpdf method, generally yields the worst predictions of the study, except in the case of the standard deviation where it outperforms the simple linear regression. This illustrates the importance of using the correct bivariate distribution for the construction of the conditional probability density function. The WR method, based on a nonlinear regression approach, has a very similar, though slightly poorer performance than the SLR. The SLR method only yields a good prediction of the long-term mean, but fails to predict all quantities involving fluctuations. Finally, the VR technique shows a good prediction of both mean and standard deviation, as expected from Eq. (6), but is less reliable when it comes to predicting Weibull parameters and the energy density.

Some quite general conclusions can be drawn from the current study: neither the modeling approach (regression versus kernel) nor the parameter estimation scheme alone decides upon the

TABLE III. Qualitative summary of the performance of each method tested with respect to each metric.

Method	Metric				
	Mean	Standard deviation	Weibull scale factor	Weibull shape factor	Energy density
SLR	Well predicted	Generally underpredicted; prediction accuracy dependent on correlation coefficient	Underpredicted by 0%–3%	Mostly overpredicted (0%–15%). Works better if target < reference shape factor	Always strongly underpredicted
WR	Well predicted	Generally underpredicted; prediction accuracy dependent on correlation coefficient	Underpredicted by 0%–4%	Always overpredicted (about 10%). No dependence on shape factors.	Always strongly underpredicted
VR	Well predicted	Fairly well predicted. Works best if target < reference shape factor	Underpredicted by 0%–3%	Underprediction by 0%–12%. Works better if target < reference shape factor.	Mostly underpredicted
SLRpdf	Overpredicted by 1%–12%	Underpredicted, but better predictions than from SLR and WR.	Overpredicted by 1%–12%	Always overpredicted (5% and more). Works better if target < reference shape factor. Works best if shape factors $\sim 3$ .	Strongly overpredicted.
Wpdf	Well predicted	Well predicted	Well predicted	Well predicted	Well predicted

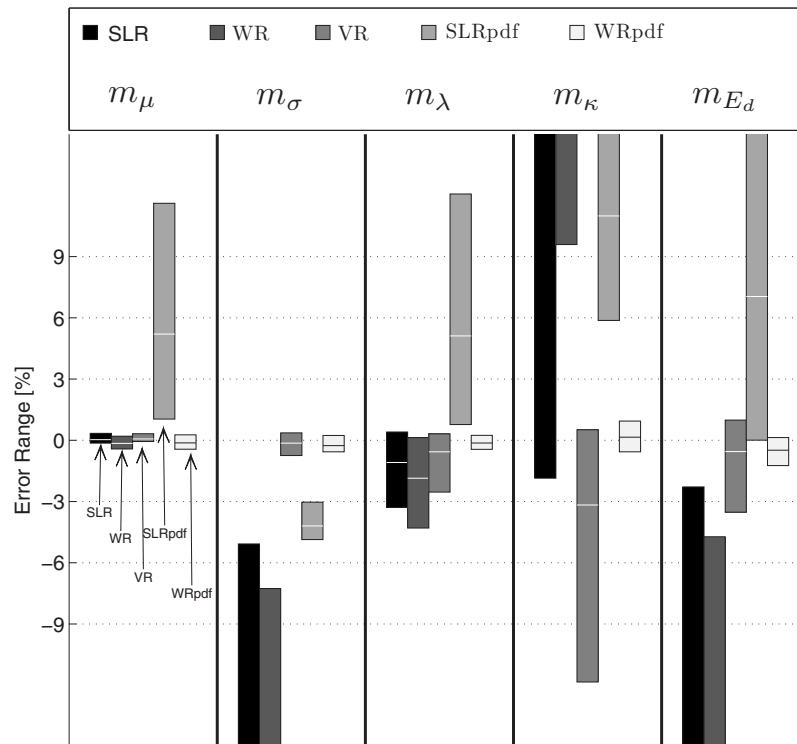


FIG. 10. Summary of the results for each of the five metrics and each of the five MCP methods studied covering a range of both correlation coefficients and shape factor ratios.  $\lambda_x=\lambda_y=7.5$ , concurrency period=9500,  $\kappa_x/\kappa_y=1.3, 1.6, 1.9, 2.2, 2.5, 2.8$  with  $\kappa_x=2.4$ .

suitability of an MCP method, but rather the combination of the two. Only the Wpdf method, which uses both the kernel approach and the conditional probability density function which most resembles the natural distribution of the data, provides an accurate prediction of all relevant output parameters or metrics for all combinations of the input parameters. The method using Weibull parameters but limited to the regression approach (the WR method) does not yield any significant improvement over the SLR. The model using Gaussian kernels consistent with the SLR parameters (SLRpdf) yields predictions that are generally even worse than those of the simple linear regression, with the exception of the prediction of the standard deviation and the special case of high and equal shape factors ( $\kappa_x=\kappa_y\sim 3$ ) where the Weibull distribution resembles a normal distribution.

As in the study by Rogers *et al.*,<sup>1</sup> the VR method predicts all output parameters surprisingly well, performing as well as the Wpdf method except for conditions where the reference and the target shape factors are different. The fact that it requires only the average wind speeds and standard deviations at both sites as input parameters reduces its computational requirements and allows for an easy implementation on a spreadsheet. The Wpdf method, on the other hand, stands out from all methods studied as far as its prediction accuracy is concerned and can be run easily on an off-the-shelf PC, albeit requiring a little more programming effort than the simple VR method.

## ACKNOWLEDGMENTS

The authors thank Instituto Tecnológico y de Estudios Superiores de Monterrey for the support through the Research Chair for Wind Energy (Account No. CAT158).

- <sup>1</sup> A. L. Rogers, J. W. Rogers, and J. F. Manwell, *J. Wind Eng. Ind. Aerodyn.* **93**, 243 (2005).
- <sup>2</sup> M. Andersen, "A review of measure-correlate-predict techniques," Technical Report No. 01327R00022 (Renewable Energy Systems, 2004).
- <sup>3</sup> A. Derrick, Proceedings of the European Wind Energy Conference (EWEC), 1993.
- <sup>4</sup> W. H. Press, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, England, 1992).
- <sup>5</sup> L. Madsen, H. Joensen, and A. Landberg, Proceedings of the EWEC, 1999.
- <sup>6</sup> L. Mortensen, N. G. Barthelmie R. J. Joensen, A. Nielsen, and M. Landberg, Proceedings of the EWEA, 2001.
- <sup>7</sup> J. C. Woods and S. J. Watson, *J. Wind Eng. Ind. Aerodyn.* **66**, 85 (1997).
- <sup>8</sup> M. Riedel and V. Strack, Proceedings of the EWEA, 2001.
- <sup>9</sup> A. Mortimer, Proceedings of the BWEA, 1994.
- <sup>10</sup> P. A. Jiménez, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, *J. Atmos. Ocean. Technol.* **27**, 1101 (2010).
- <sup>11</sup> A. Romo, J. Amezcua, and O. Probst, *Proceedings of Wind Power 2009* (AWEA, Chicago, 2009).
- <sup>12</sup> S. E. Tuller and A. C. Brett, *J. Clim. Appl. Meteorol.* **23**, 124 (1984).
- <sup>13</sup> G. C. Alexandropoulos, N. C. Sagias, and K. Berberidis, *IEEE Antennas Wireless Propag. Lett.* **6**, 93 (2007).
- <sup>14</sup> Y. A. Chau and K. Y.-T. Huang, *Electron. Lett.* **43**, 138 (2007).
- <sup>15</sup> D. W. Matolak, I. Sen, and W. Xiong, *IET Communications* **2**, 523 (2008).
- <sup>16</sup> L. M. Klyatis, O. I. Teskin, and J. W. Fulton, "Multi-variate Weibull model for predicting system-reliability, from testing results of the components," pp. 144–149, 2000.
- <sup>17</sup> N. C. Sagias and G. K. Karagiannidis, *IEEE Trans. Inf. Theory* **51**, 3608 (2005).
- <sup>18</sup> R. A. Johnson, J. W. Evans, and D. W. Green, "Some bivariate distributions for modeling the strength properties of lumber," United States Department of Agriculture: Forest Service, 1999.
- <sup>19</sup> C. P. Chan, J. R. Stalker, A. Edelman, and S. R. Connors AWEA 2010 Conference Proceedings, 2010.
- <sup>20</sup> D. Oakes and A. K. Manatunga, *Biometrika* **79**, 827 (1992).
- <sup>21</sup> C. Cox and G. Ma, *Biometrics* **51**, 142 (1995).
- <sup>22</sup> H. Lütkepohl, *Introduction to Multiple Time Series Analysis*, 2nd ed. (Springer-Verlag, Berlin, 1993).
- <sup>23</sup> G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Wiley Series in Probability and Statistics, 4th ed. (Wiley, Hoboken, NJ, 2008).
- <sup>24</sup> L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, New York, 1986).
- <sup>25</sup> W.-H. Chung, R. E. Hudson, and K. Yao, *IEEE Trans. Commun.* **57**, 3481 (2009).
- <sup>26</sup> T. Burton, *Wind Energy Handbook* (NetLibrary, Inc., Chichester, 2001).
- <sup>27</sup> G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. (Duxbury/Thomson Learning, Pacific Grove, CA, 2002).